On macromolecular refinement at subatomic resolution with interatomic scatterers

Pavel V. Afonine[*$], Ralf W. Grosse-Kunstleve[$], Paul D. Adams[$], Vladimir Y. Lunin[+], Alexandre Urzhumtsev[€]

[$]*Lawrence Berkeley National Laboratory, One Cyclotron Road, BLDG 64R0121, Berkeley, CA 94720 USA*

[+]*Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, 142290 Russia*

[€] *IGMBC, 1 rue L.Fries, 67404 Illkirch & IBMC, 15 rue R.Descartes, 67084 Strasbourg, & Faculty of Sciences, Nancy University, 54506 Vandoeuvre-lès-Nancy, France*

Email: PAfonine@lbl.gov

Synopsis

Modelling deformation electron density using inter-atomic scatters is simpler than multipolar methods, and produces comparable results at subatomic resolution and can be easily applied to macromolecules.

Abstract

A study of the accurate electron density distribution in molecular crystals at subatomic resolution, better than ~1.0 Å, requires more detailed models than those based on independent spherical atoms. A tool conventionally used in small-molecule crystallography is the

multipolar model. Even at upper resolution limits of 0.8-1.0 Å, the number of experimental data is insufficient for the full multipolar model refinement. As an alternative, a simpler model composed of conventional independent spherical atoms augmented by additional scatterers to model bonding effects has been proposed. Refinement of these mixed models for several benchmark datasets gave results comparable in quality with results of multipolar refinement and superior of those for conventional models. Applications to several datasets of both small- and macro-molecules are shown. These refinements were performed using the general-purpose macromolecular refinement module *phenix.refine* of the *PHENIX* package.

Keywords: structure refinement, subatomic resolution, deformation density, interatomic scatterers, PHENIX.

1. Introduction

The growing number of macromolecular crystals diffracting to subatomic resolution (53 models in 2003; currently 270) requires the development of appropriate methods and software to best model them. The new information obtained from such macromolecular studies has been discussed in a number of articles (see for example reviews Dauter *et al*., 1995, 1997; Vrielenk & Sampson, 2003; Petrova & Podjarny, 2004, and numerous references therein). Afonine *et al*. (2004) have shown that information about the density deformation of individual atoms can be extracted from macromolecular data at resolutions of 0.9 Å or better. As a consequence, conventional models for macromolecular structures where the electron density of the molecule is a simple sum of contributions from spherical atoms smeared by individual anisotropic displacements are incomplete and provide inaccurate values for ADPs (Atomic

Displacement Parameters). Following previous publications we refer to these models as IAM (Independent Atom Models).

Model refinement of small molecules at subatomic resolution largely uses the multipolar formalism of Hansen & Coppens (1978). For these models the electron density is a sum of atomic contributions where the density is no longer spherical but depends on the chemical environment. Such a non-spherical distribution is described by a linear combination of spherical harmonics (Hansen & Coppens, 1978). Refinement of parameters of multipolar models is monitored mainly by decrease of the crystallographic *R*-factor, improvement of the residual Fourier syntheses, the rigid-bond test (*RBT*) and other characteristics.

Lecomte and co-workers have reported a number of multipolar refinements of amino- and nucleic acids to determine a database of multipole parameters and described several cases of polypeptide and protein refinement using this database (for a review see Jelsch *et al*., 2005). Recently the group of Coppens (Volkov *et al*., 2007; VC07 in what follows) also reported an application of the multipolar refinement to polypeptides, but using their own database for multipolar parameters.

VC07 concluded that the applicability of multipolar models in macromolecular studies "is in general not warranted, unless exceptionally high-resolution data of ~0.6 Å or better with satisfactory completeness" are available. Also it was stated that "for macromolecular crystal such data are generally not available, (…) the number of reflections is not sufficient". A possible solution to overcome this obstacle is a direct transfer of library parameters without their refinement as discussed by Brock *et al*. (1991), Pichon-Pesme *et al*. (1995), Jelsch *et al*. (1998), Dittrich *et al*. (2005), VC07 and Zarychta *et al*. (2007). However, since the quality of macromolecular X-ray data is generally lower than for small-molecule crystals, an alternative solution is to introduce a model of intermediate complexity, more detailed than IAM but simpler than a multipolar one (Afonine *et al*., 2004). A possible approach is to complete the

IAM with spherical scatterers between the atoms (IAS, interatomic scatterers). It should be noted that the use of the IAM-IAS model is much more runtime efficient and can be straightforwardly implemented in macromolecular crystallographic packages. We here use 'IAS' instead of the previous name 'DBE', Dummy Bond Electron model (Afonine *et al*., 2004), reflecting better the features of the model.

In this paper we compare the results obtained with different types of electron density models for several benchmark data sets. The implementation of IAS modelling into the general purposes crystallographic program suite *PHENIX* (Adams *et al*., 2002) has allowed the corresponding refinements with *phenix.refine* (Afonine *et al*., 2005) to be performed quickly and in a fully automated fashion.

2. Comparative refinement at subatomic resolution

The modelling of structures at a subatomic resolution with multipolar models takes into account the delocalization of electron density from atomic centres due to formation of interatomic bonds. The IAM-IAS model (Afonine *et al*., 2004) instead treats this delocalized density as spherical Gaussian scatterers located at the centroid of the delocalized density and keeps conventional spherical atoms unchanged. The multipolar model requires that existing IAM atoms be replaced, while the IAS models complete them by specifically constructed scatterers. Also, the IAM-IAS model may be gradually extended once the new features become visible. Some details of the construction and refinement of IAM-IAS models, and development of the corresponding library of parameters were originally outlined by Afonine *et al*. (2004). The current tests were aimed to demonstrate that IAM-IAS models can improve conventional IAM models by lowering R and R_{free} factors, correcting ADP parameters and producing clearer residual maps to the same degree as multipolar models yet are significantly

simpler to work with. In this short communication we have no possibility to discuss applications other than map improvement (see for example Afonine *et al*., 2002). By the same reason the complete methodology and implementation details of IAS in *PHENIX* including choice of the refinement targets, role of data completeness and the efficient resolution, will be discussed separately in a full length paper (Afonine *et al*., in preparation).

To estimate the quality of IAM-IAS models, we built and refined such models for YGG and P2A4 (Table 1) for which a comparative refinement has been reported by VC07. Similarly to VC07, refinement was performed at two different resolutions. The highest available resolution (0.44 Å and 0.37 Å, respectively; for YGG the data completeness is below 50% at a resolution higher than 0.57 Å) was considered as a 'high resolution' where the data-to-parameter ratio is high enough even for the use of a multipolar model, and a resolution of 0.80 Å was the 'low resolution' where this ratio becomes too low. In addition to the standard *R*-factor and rigid-bond test (Hirshfeld, 1976), R_{free} (Brünger, 1992) was used as a refinement quality indicator.

Unfortunately, the YGG and P2A4 models have been refined previously against the full set of data (in fact, the set selected with $I > 3\sigma(I)$ that is not explicitly stated in VC07) making the conventional R_{free}-analysis biased. Therefore when performing the IAS-refinements, we only note that R_{free} is lower than the corresponding values for the refined IAM models.

The IAM-IAS models were generated and refined completely automatically in *PHENIX*. Table 2 shows principal refinement information. All stereochemical and ADP restraints on atomic parameters were removed for both small- and macromolecules used in this study (Dauter *et al*., 1997; Schmidt *et al*., 2003; Petrova *et al*., 2006). Since the starting models were previously refined to a high quality no stereochemical distortions due to the unrestrained refinement were observed. Decrease of the R_{free} shows that refinement of IAS did

not overfit the experimental data and indeed improved the models. When refining at 'low resolution', the ADP values obtained with the IAS are smaller than those from the refinement of corresponding IAMs. Based on previous works (Afonine *et al*., 2004; Petrova *et al*., 2006), we believe that they are closer to the correct values of ADP, which otherwise will tend to increase to model the deformation density along the bonds (Coppens, 1967; Dunitz & Seiler, 1973). The rigid-bond test also confirms that the introduction of IAS improved the model. In fact, the IAS-refinement with the Maximum Likelihood target (Lunin *et al*., 2002; to our knowledge never previously applied in this context), improved the models further as measured by the rigid-bond test, however that analysis is beyond the scope of this paper. For 'high-resolution refinement', mean ADP values are similar with and without IAS, as noted previously by Afonine *et al*. (2004). This indicates that the highest resolution data contain enough information to deconvolute the deformation density and atomic uncertainty effects and to estimate ADP correctly even without IAS.

Fig. 1 illustrates the improvement of the difference F_{obs}-F_{calc} maps reducing the residual peaks to the same level as for multipolar models (compare with Figs. 2 and 3 in VC07). Overall, for the whole set of monitoring parameters the results show the comparable quality of the IAM-IAS and multipolar models despite the simplicity of the former.

Several macromolecular structures were used as another benchmark (Table 1). Previously, refinement at subatomic resolution using multipolar models has been reported for crambin (Fernandez-Serra *et al*., 2000; 0.54 Å), trypsin (Schmidt *et al*., 2003; 0.80 Å), phospholipase (Liu *et al*., 2003; 0.80 Å; for the resolution higher than 0.86 Å the data completeness is below 50%) and scorpion toxin (Housset *et al*., 2000; 0.96 Å). The corresponding models were extracted from the PDB (Bernstein *et al*., 1977; Berman *et al*., 2000). Unfortunately, the models available in PDB did not allow reproducing exactly the results reported making comparative analysis of the IAS-refinement also impossible. In

particular, this completely excluded the crambin data from our tests. To complete the picture at higher resolution, we additionally performed an IAS-refinement at 0.62 Å of the antifreeze protein RD1 (Ko *et al*., 2003). Table 2 summarizes the results of refinement of these models. For all cases the residual maps became much clearer. In particular, this map improvement highlighted the double conformation of the S-S bond for the phospholipase and trypsin structures that are otherwise hidden in the noise and identified two ions previously interpreted as waters (Fig. 1c illustrates this for RD1).

In all cases, the full round of completely automated IAS model building and IAS-IAM refinement, with no manual intervention, took from a few minutes to an 1 hour on a modern Linux computer. For all protein refinements, completing IAM by IAS decreases the *R*-factors; also the R_{free}-factors are lower for IAS-IAM than for IAM. The *RBT* value systematically decreases after the introduction of IAS. We observe that the mean ADP slightly increased for the scorpion toxin data, which may indicate that the resolution (0.96 Å) approaches the limit for the use of the IAM-IAS method.

3. Conclusion and Acknowledgments

Currently multipolar modelling is the most precise and powerful tool for crystallographic studies at subatomic resolution when the crystals diffract to ultra-high resolution, about 0.6 Å or higher, and the data-to-parameter ratio justifies refinement of the model parameters. At a resolution near 0.8-0.9 Å, more common for macromolecular crystals at sub-angstrom resolutions, multipolar modelling typically requires too many parameters to be refined. As an alternative to the multipolar method, IAM-IAS models may be used, where IAM atoms are augmented by small interatomic scatterers. This approach makes model

building and refinement a very transparent and easily monitored procedure. Results of automated refinement of such models for both small- and macromolecules at subatomic resolution confirm the efficiency of these models, both in terms of model quality and CPU resources required. The tests show that these models can be used even at ultra-high resolution producing results comparable with those obtained with multipolar models.

This work was supported in part by the US Department of Energy under Contract No. DE-AC03-76SF00098 and a grant to PDA from NIH/NIGMS (1P01GM063210). VYL was supported by RFBR grants 05-01-22002_CNRS-a and 07-07-00313-A. All results presented are based on the *CCI Apps* source code bundle with the version tag 2007_08_10_0051. *PyMOL* (DeLano, 2002) was used to present the maps and structures.

References

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.-W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K. & Terwilliger, T.C. (2002). *Acta Cryst.*, **D58**, 1848-1854.

Afonine, P.V., Pichon-Pesme, V., Muzet, N., Jelsch, C., Lecomte, C. & Urzhumtsev, A. (2002). *CCP4 Newsletter on Protein Crystallography*. **41**, electronic publication: http://www.ccp4.ac.uk/newsletter41/00_contents.html

Afonine, P., Lunin, V.Y., Muzet, N. & Urzhumtsev, A. (2004). *Acta Cryst.*, **D60**, 260-274.

Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl.* **42**, contribution 8.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J.Mol.Biol*., **112**, 535-542.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000). *Nucleic Acids Research*. **28**, 235-242.

Brock, C.P., Dunitz, J.D. & Hirshfeld, F.L. (1991). *Acta Cryst*., **B47**, 789-797.

Brünger, A. (1992). *Nature*, **355**, 472-475.

Coppens, P. (1967). *Science*, **158**, 1577-1579.

Dauter, Z., Lamzin, V.S. & Wilson, K.S. (1995). *Curr.Opinion in Struct.Biol.*, **5**, 784-790.

Dauter, Z., Lamzin, V.S. & Wilson, K.S. (1997). *Curr.Opinion in Struct.Biol.*, **7**, 681-688.

DeLano, W.L. (2002). *The PyMOL Molecular Graphics System,* DeLano Scientific, San Carlos, CA, USA. *http://www.pymol.org*

Dittrich, B., Kortisánszky, T., Grosche, M., Scherer, W., Flaig, R., Wagner, A., Krane, H.G., Kessler, H., Riemer, C., Schreurs, A.M.M. & Luger, P. (2002). *Acta Cryst*., **B58**, 721-727.

Dittrich, B., Hübschle, C.B., Messerschmidt, M., Kalinowski, R., Girnt, D. & Luger, P. (2005). *Acta Cryst*., **A61**, 314-320.

Dunitz, J.D. & Seiler, P. (1973). *Acta Cryst.,* **B29**, 589-595.

Fernandez-Serra, M.V., Junquera, J., Jelsch, C., Lecomte, C. & Artacho, E. (2000). *Solid State Communications,* **116**, 395-400.

Hansen, N.K. & Coppens, P. (1978). *Acta Cryst.*, **A34**, 909-921.

Hirshfeld, F.L. (1976). *Acta Cryst.*, **A32**, 239-244.

Housset, D., Benabicha, F., Pichon-Pesme, V., Jelsch, C., Maierhofer, A., David, S., Fontecilla-Camps, J.C. & Lecomte, C. (2000). *Acta Cryst.,* **D56**, 151-160.

Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Acta Cryst.,* **D54**, 1306-1318.

Jelsch, C., Teeter, M.M., Lamzin, V.S., Pichon-Pesme, V., Blessing, R.H. & Lecomte, C. (2000). *Proc.Natl.Acad.Sci. USA*, **97**, 31731-3176.

Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J.Appl. Cryst.*, **38**, 38-54.

Ko, T.-P., Robinson, H., Gao, Y.-G., Cheng, C.-H.C., DeVries, A.L. & Wang A.H.-J. (2003). *Biophys. J.,* **84**, 1228 –1237.

Lunin, V.Y., Afonine, P.V. & Urzhumtsev, A.G. (2002). *Acta Cryst*. **A58**, 270-282.

Liu, Q., Huang, Q., Teng, M., Weeks, C.M., Jelsch, C., Zhang, R. & Niu, L. (2003). *J.Biol.Chem.,* **278**, 41400-41408.

Petrova, T.E. & Podjarny, A.D. (2004). *Rep.Prog.Phys.,* **67**, 1565-1605

Petrova, T.E., Ginell, S., Mitschler, A., Hazemann, I., Schneider, T., Cousido, A., Lunin, V.Y., Joachimiak, A. & Podjarny, A.D. (2006). *Acta Cryst*., **D62**, 1535-1544.

Pichon-Pesme, V., Lecomte, C. & Lachekar, H. (1995). *J.Phys.Chem.,* **99**, 6242-6250.

Schmidt, A., Jelsch, C., Østergaard, P., Rypniewsli, W. & Lamzin, V.S. (2003). *J.Biol.Chem.,* **278**, 43357-43362.

Volkov, A., Messerschmidt, M. & Coppens, P. (2007). *Acta Cryst*. **D63**, 160-170.

Vrielink, A. & Sampson, N. (2003). *Curr.Op.Str.Biol.,* **13**, 709-715

Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. & Jelsch, C. (2007). *Acta Cryst.,* **A63**, 108-125.

Fig. 1. Residual Fourier maps calculated in absolute scale. IAS are shown by small spheres, in magenta (IAS with positive occupancy) and in tint (IAS with negative occupancy). **(first and second rows).** Maps at 0.43 Å resolution for YGG. Left and middle: IAM phased maps, right: IAM-IAS phased maps. Contour colours are: +0.20 eÅ^{-3} (marine), +0.10 eÅ^{-3} (cyan), -0.10 eÅ^{-3} (yellow), -0.20 eÅ^{-3} (red). Views are similar to those in Figs. 2-3 of VC07. **(third row).** Maps at 0.62 Å resolution for the antifreeze protein RD1. Left and middle: IAM phased maps shown at cut-off levels 0.40 eÅ^{-3} (green) and 0.25 eÅ^{-3} (light blue); right: IAM-IAS phased map shown at a cut-off level 0.25 eÅ^{-3} (light blue). The SO_4 ion inserted instead of previously located water nicely fits the residual density (shown in tint).

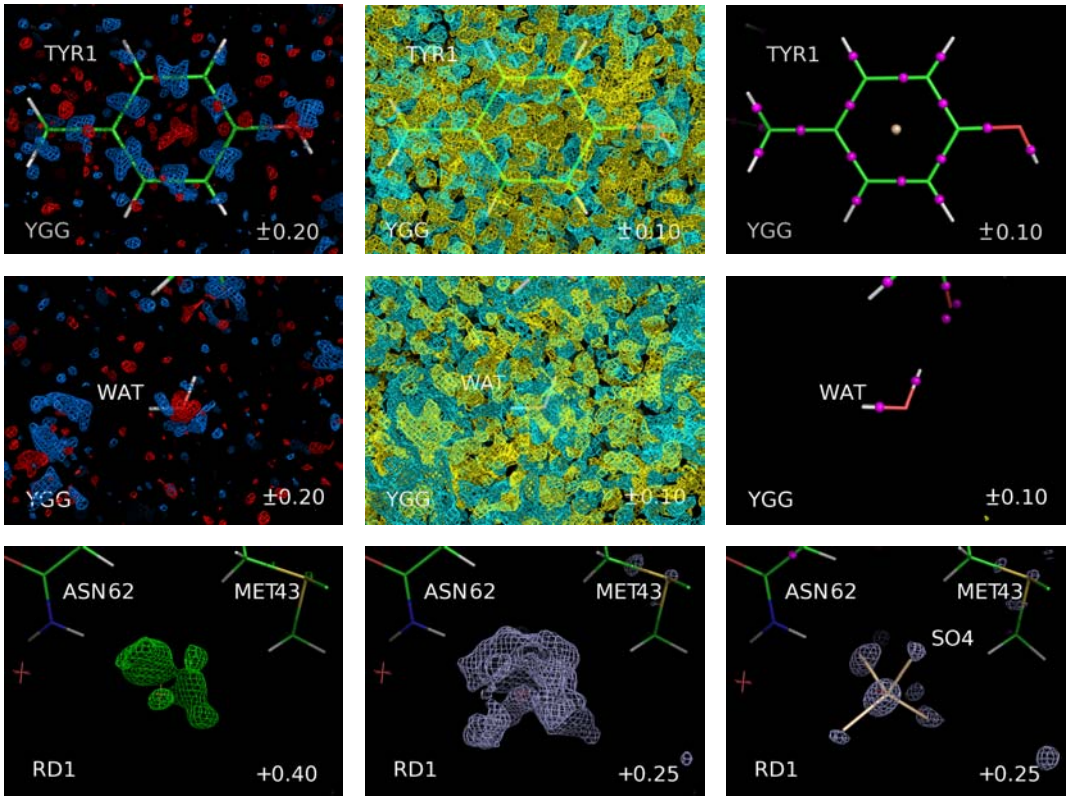

Table 1. Data used for refinements. N_{nonH}, N_{H} and N_{IAS} give the number of non-hydrogen, hydrogen and IAS atoms in corresponding models. d_{high} is the highest resolution for the data set, N_{high} is the corresponding number of reflections. N_{low} is the number of reflections for the data sets truncated to lower resolution (d_{low}= 0.80 Å, for YGG and P2A4 only). References for multipolar refinement are: HL00 = Housset *et al*., 2000; JL00 = Jelsch *et al*., 2000; KW03 = Ko *et al*., 2003; LN03 = Liu *et al*., 2003; SL03 = Schmidt *et al*., 2003; VC07 = Volkov *et al*., 2007.

Molecule (reference)	Space group Unit cell (in Å and °)	N_{nonH}	N_{H}	N_{IAS}	d_{high} (in Å)	N_{high}	N_{low}
YGG (VC07)	$P2_12_12_1$ 7.98, 9.54, 18.32	22	19	39	0.43	4766	1358
P2A4 (VC07)	$P2_12_12_1$ 10.13, 12.50, 19.50	35	36	71	0.37	21475	2513
antifreeze protein (KW03)	$P2_12_12_1$ 32.50, 39.50, 44.64	650	518	367	0.62	118501	-
trypsin (SL03)	$P1$ 32.87, 37.02, 39.78, 102.89, 104.59, 102.37	2231	1515	1362	0.80	163918	-
phospholipase (LN03)	$C2$ 44.73, 59.09, 45.31, 90.00, 117.43, 90.00	1324	956	679	0.80	77695	-
scorpion toxin (HL00)	$P2_12_12_1$ 45.90, 40.70, 30.10	647	441	335	0.96	31001	-

Table 2. Comparative statistics for refinement of IAS and multipolar models. Models marked by [(a)] were refined by VC07 (Volkov *et al*., 2007); corresponding numbers are cited from there. M_t and M_r stand for multipolar models with transferred and refined parameters (refinements '3' and '5' in VC07). $\langle B_{\mathrm{nonH}} \rangle$ stand for the mean value of the equivalent isotropic ADP, in Å^2, calculated for non-hydrogen atoms; *RBT* is the rigid-bond-test value, in 10^4 Å^2 (the same as *DMSDA*, differences in mean-squared displacement amplitudes, in VC07). R_{work} and R_{free} stand for standard crystallographic R and R_{free} factors between experimental F^{obs} and model-based calculated structure factor magnitudes F^{model} (Afonine *et al*., 2005) as $\left[\sum_{\mathbf{s}} \left| F_{\mathbf{s}}^{obs} - kF_{\mathbf{s}}^{model} \right| \right] \left[\sum_{\mathbf{s}} F_{\mathbf{s}}^{obs}\right]^{-1}$. [(b)]An estimate; obtained if the same set of parameters were used for refinement at 'high' resolution. [(c)]For multipolar refinement a number of parameters were fixed or linked by constraints; N_{par} is the number of parameters at each step and does not include the number of parameters refined previously. Differently from VC07, for the current project the ratio N_{data}/N_{par} is calculated for the total number of refined parameters even when at each particular moment only a subset of them were refined; a direct comparison of this information with that reported in VC07 is not straightforward.

data set	model	N_{data}/N_{par} [(c)]	R_{work}	R_{free}	$\langle B_{\mathrm{nonH}} \rangle$	*RBT*
YGG-low resol.	IAM[(a)]	4.9	2.16	-	-	17.76
	M_t[(a)]	6.2	1.22	-	-	12.85
	IAM	6.2	2.35	2.62	1.23	18.99
	IAS	4.0	1.57	2.00	1.05	12.23
YGG-high resol.	IAM[(a)]	17.3	4.51	-	-	8.77
	M_t[(a)]	21.9	3.66	-	-	7.38
	M_r[(a)]	10.6 [(b)]	3.42	-	-	6.38
	IAM	21.9	4.57	4.72	1.04	8.62
	IAS	14.2	3.75	4.06	1.07	7.68
P2A4-low resol.	IAM[(a)]	5.5	2.98	-	-	15.64
	M_t[(a)]	7.1	1.84	-	-	7.09
	IAM	7.1	3.51	3.79	1.24	20.77
	IAS	4.5	2.45	3.27	1.07	16.77
P2A4-high resol.	IAM[(a)]	46.7	3.44	-	-	3.67
	M_t[(a)]	61.0	2.67	-	-	2.65
	M_r[(a)]	43.6 [(b)]	2.53	-	-	3.09
	IAM	61.1	3.72	3.63	1.14	3.66
	IAS	38.1	3.06	3.23	1.14	4.79
antifreeze protein	IAM	18.6	12.77	15.37	7.84	208.4
	IAS	14.3	11.76	14.44	7.40	195.7
trypsin	IAM	7.6	10.30	13.79	5.79	149.3
	IAS	5.8	9.19	13.35	5.52	126.0
phospholipase	IAM	6.0	8.99	12.80	9.88	250.6
	IAS	4.7	8.31	12.64	9.11	213.5
scorpion toxin	IAM	4.9	9.40	15.47	10.30	365.8
	IAS	3.9	8.78	15.23	10.42	363.1